Digital Twins in Geospatial Research

Report of the SPARC Workshop held at ASU's Tempe campus on February 27–28 2023

Background

For the past five years ASU's Spatial Analysis Research Center (SPARC) has organized annual workshops on topics at the cutting edge of spatial analysis and geographic information science (https://sgsup.asu.edu/sparc/workshops). Each in-person workshop has involved about 30 participants, and has included keynote speeches, lightning talks, plenary and group discussions, and opportunities for social interaction. The 2019 workshop on "Replicability and Reproducibility" was followed by a 2020 workshop on "Scale and Spatial Analytics." The onset of the pandemic shortly after the 2020 workshop required that the 2021 and 2022 workshops be held online in a much shorter format as webinars. SPARC was able to take advantage of the series of webinars being organized by the American Association of Geographers and its GeoEthics project, offering one on "Ethical Spatial Analytics" in 2021 and one titled "Does GeoAI Offer an Ethical Future for Spatial Analytics?" in 2022. Both of these workshops reached several hundred people. 2023 marked a return to in-person workshops, and "Digital Twins in Geospatial Research" was chosen as a suitably cutting-edge topic. This report provides an overview of the workshop and its major findings.

The 2023 workshop was organized by SPARC faculty members Peter Kedron, Wenwen Li, Daoqin Tong, Dylan Connor, Stewart Fotheringham, Amy Frazier, and Michael Goodchild. Director of ASU Geospatial Research Services Shea Lemar also contributed to the logistical planning of the workshop. Support was provided by staff members and students of the School of Geographical Sciences and Urban Planning (SGSUP) that are listed later in this report. Funding for the workshop, including travel support for most of the participants, was provided by a grant from Esri and from SPARC. The organizers thank the sponsors and the supporting staff for their excellent work in making the workshop a success.

Motivation

A digital twin can be defined as "A virtual representation of the real world, including physical objects, processes, relationships, and behaviors" (Esri). As such it allows experiments to be carried out on the digital twin rather than on reality, and allows predictions to be made about the behavior of the real system without incurring the costs and side effects of implementing such intervention. Digital twins can offer access to many types of data associated with a system and to the processes that govern the system's dynamic behavior, and their emphasis on fine spatial resolution can support very compelling visualizations. Digital twins have been attracting attention in recent years in a host of fields ranging from manufacturing and urban planning to oceanography and ecology.

Digital twins overlap with geographic information systems whenever the system to be twinned exists in the geographic domain and at geographic scales—that is, on or near the Earth's surface and oceans and at resolutions from 1cm up to 10km. In this context Esri goes so far as to assert that "GIS is foundational for any digital twin," and in the remainder of this report the GIS foundation will be assumed. The concept is clearly related to the Digital Earth that was the subject of a speech by Vice-President Gore in

1998 and has been the theme of much literature, a series of international conferences, an international journal, and a recent manual (Guo, Goodchild, and Annoni, 2020). However, the Digital Earth as conceived by Gore and implemented in Google Earth falls somewhat short of the current vision of a digital twin in lacking representations of processes and thus the opportunity for interventions and predictions.

The need for a specialist meeting

Much of the energy behind digital twins derives from their value in prediction and assessment, and the academic community, with its traditional emphasis on explanation and understanding, has paid them little attention to date. In late 2022 NASA held a highly over-subscribed workshop on digital twins, with strong participation from the private sector and NASA-supported research groups. The membership of the Digital Twin Consortium includes a handful of academic institutions (Monash University, University of Melbourne, University of Alaska at Anchorage, the University of Maryland's Center for Environmental Energy Engineering, and the Universitat Politecnica Catalunya are currently listed as members), and other digital twin projects have roots and links in academia (such as the Gemini Project at the University of Cambridge, and the DITTO Project, which promotes the construction of digital twins of the oceans). Indeed, digital twins have garnered sufficient attention that the National Academy of Sciences, Engineering, and Medicine appointed an ad hoc committee on the topic and held three information-gathering workshops in early 2023. While this effort will certainly move the topic forward on many disciplinary research agendas, none of these workshops focused on key geographic questions surrounding the development and use of digital twins.

As such, it is time for a more comprehensive and dedicated academic and geographic look at the concept; one that is able to ask the kinds of questions that are characteristic of academia:

- about the reliability and validity of predictions from digital twins,
- about the deeper technical issues of digital twin design,
- about the ethical and societal implications of digital twins, and
- about educating and training the next generation of digital twin developers and users.

These four themes will be revisited in later sections of this report.

Underlying the recent interest in digital twins is the question "Why now?" There is nothing new in the concept of simulating a real system using a digital representation of knowledge about the system and the processes that operate within it; these ideas occur in many domains, and date from the earliest days of computing. The technology to support such simulations has been advancing, along with the availability of suitable data, and the spatial and temporal resolutions of those data have been improving steadily. But no threshold has been reached that might explain the sudden surge in the use of the term "digital twin" in recent years. No digital twin can ever be identical to the system it represents, and no predictions can ever be exactly true. Biologically identical twins are never exactly the same, since their identical DNA is expressed in ways that are always modified by nurture and exposure. Thus, no digital twin will ever pass a Turing test by leaving the user unable to determine whether the response to a query came from the digital system or from the reality it attempts to represent, though digital twins in manufacturing will likely come closer than those in social and environmental contexts.

One possible answer to the "Why now?" question is specific to the geospatial case. During the workshop Sean Ahearn suggested that LiDAR and RGB-D technologies are now capable of providing strikingly realistic visualizations that users may be incapable of distinguishing from actual views, and in this narrow sense the Turing test may already have been passed. Another more general answer concerns possible changes in the culture of science and its funding mechanisms. Perhaps the general speed-up in science, and the consequent lack of time for reflection and critique, may have simply made science more susceptible to fads and viral trends than in the past.

The opening session included a ten-minute video recorded by Jack Dangermond, Esri's President. He noted the difference between the early development of digital twins, for example as tools in the design process of aircraft, and their newer applications in urban planning and environmental science. Here a digital twin should be regarded as a living system, constantly receiving and integrating new information, and an important element therefore for the software industry that provides these fields with digital infrastructure. Digital twins raise some important issues that need to be discussed and explored through workshops like this one, in areas such as individual privacy, surveillance, and tracking; uncertainty in how data and models represent reality; their role as infrastructure for open scientific discovery, monitoring, and forecasting; and the societal and political implications of engaging the citizenry through the technologies of smart cities and digital twins.

Keynote speeches

A total of eight keynote speeches lasting 30 minutes each were given at the workshop, organized into matching pairs and spread over the two days.

Keynote Pairing I – Mike Batty and Maryann Feldman

The first pair focused on urban and regional planning, a key area of application of digital twins with the potential to visualize and predict the effects of planning decisions. **Mike Batty** of University College London, a widely recognized expert on the use of quantitative methods in urban planning and early author on the potential for digital twins in the geospatial realm, gave the first keynote. He showed numerous examples of the work of his Centre for Advanced Spatial Analysis in modeling urban systems at a variety of spatial scales, ranging from transport congestion in London to the impacts of UK-scale decisions such as the high-speed rail link from London to the North of England. He saw major issues emerging in the variety of models that can be applied to a single system, and the range of results that they might produce. The same problem arises in simulations of the atmosphere, since there are now dozens of models that might be employed to predict the effects of climate change, producing an *ensemble* of results. To quote British statistician G.E.P. Box, writing long before the concept of digital twins became popular, "All models are wrong but some are useful." Mike drew attention to the problem that arises when decision-makers must somehow choose between the predictions of several (wrong) models.

The second keynote in this pair was given by economist **Maryann Feldman** of ASU. Her work concerns the impacts of policy decisions, and digital twins often provide a very effective approach. Why is it, for example, that the same policy decision may impact different locations in very different ways? In other words, our understanding of urban environments may not replicate well from one city to another. That leads in turn to a key question for planners: how do we turn around the fortune of places? The use of digital twins in this context has a long history, dating back to early tools such as SimCity that provided a visual environment in which the impacts of interventions could be explored, though admittedly not always quantitatively. She discussed several examples, such as the comparative impact of pharmacological research in Boston and Washington DC and highway development in New York City and Toronto.

In the discussion following this first pair of keynotes several members of the audience picked up on the issue of uncertainty. Maryann offered a response based on scale, arguing for its importance in modeling cities. Boston, Washington DC, and New York may be simply too large to simulate, but analysis of so-called *micropolitan* areas of around 50,000 population may be more productive. Mike Batty asked whether the models being used in digital twins were appropriately structured to handle uncertainty, drawing on the literature that argues that in a geospatial context it is not adequate to add simple confidence intervals to inputs, given the importance of spatial dependence. Shaowen Wang referred to the terms “verification” and “validation,” which imply a binary distinction between truth and falsehood, whereas in reality uncertainty varies over a wide range and is never zero.

Keynote Pairing II – Kelly Easterday and Shaowen Wang

The second pair of keynotes focused on application domains, contrasting the use of digital twins in a major conservation preserve with a systematic approach to addressing the application domains of geospatial technology. **Kelly Easterday** gave the first presentation as the director of the digital twin project at the Jack and Laura Dangermond Preserve, an area of some 24,000 acres west of Santa Barbara that is managed by The Nature Conservancy. It occupies an ecologically unique location that is at the collision or convergence between the comparatively warm east-west-trending coast of Southern California and the colder north-south-trending coast around the bend represented by Point Conception. She began by commenting on a recent review of the digital twin literature and the comparative lack of digital twin projects aimed at conservation and environmental management.

The Preserve handles a large number of individual and group projects that make use of its space to investigate various ecological questions. Key however is the need to acquire and aggregate the data and models used by each individual and group into an open, accessible, and integrated resource that can be leveraged for additional projects. In that context, then, the digital twin is a data and model management solution for a multidisciplinary research enterprise.

The digital twin project at the Dangermond Preserve has four goals. First, it aims to break out of the traditional project-based model of conservation management by anticipating and managing data for the long term. Second, it seeks to accelerate the process of conservation by providing excellent digital services and support for analysis, modeling, and visualization. Third, the project aims to scale the use of its services to connect pure science and understanding to policy and decision-making. Finally, it honors the principles of open science and all that implies for documentation, replicability, and access.

Data, models, and tools are essential elements of any digital twin project, and all three require successful support in the form of infrastructure, management dashboards, and tools for story-telling. Sensor networks and communications technology are used on the Preserve to achieve near-real-time access to data. With a large community of researchers there is a clear need for people-management infrastructure, to coordinate knowledge of what each individual or group is doing in real time and to track each vehicle.

Kelly focused in detail on one major data collection project in the Jalama basin that addresses fresh-water resources. It includes base mapping, sensors in wells, and chemical sampling of water to build a comprehensive model of water budgets in the basin. Monitoring of flows in the basin is also an important component in any plan to restore the historic steelhead population.

The second keynote of the pair was given by **Shaowen Wang**, professor at the University of Illinois at Urbana-Champaign and principal investigator of the NSF-supported I-GUIDE project. He talked of the rapid evolution of technology, and its growing support for the dream of a digital twin. Shaowen emphasized how geographic scale is of critical importance to the context of digital twins. The Jalama watershed project described by Kelly covers only a small geographic area, but a key concept in I-GUIDE is telecoupling, or the tendency for what happens in one part of the globe to be directly related to events in very different parts of the world. Shaowen cited the connections that exist between farming soybeans in Illinois and in Brazil, since global markets ensure that strategies in the two areas are tightly linked, suggesting that a digital twin for soybean farming must be global.

Projects such as I-GUIDE and the Dangermond Preserve are building communities that are committed to principles of openness. Yet their scale, and the complexity of their technological base, requires that we devote more and more time to education and to preparing the next generation of the workforce. Moreover, the scale of some digital twins makes it essential that we utilize the services of high-performance computing, as represented by the ROGER installation in the case of I-GUIDE, and thus engage with the inherent complexity of such systems to ensure their accessibility to the user base.

In the discussion which followed, Clancy Wilmott and Bo Zhao raised the issue of openness and the possibility that access to some data sources will need to be restricted. Some data sources raise issues of privacy when they are sufficiently detailed to expose individuals; others concern protected species, such as the rare Dudleya plant and the nesting Peregrine falcon at the Preserve; and still others concern archaeological resources, notably the Chumash sites at the Preserve. Another issue raised during discussion concerns the role of humans in relation to digital twins. Will we ever reach the point where a digital twin knows everything about a system, or will there always be knowledge that humans possess that digital twins do not? And if the latter, then must a digital twin always fail the Turing test when queried about such knowledge?

Keynote Pairing III – Pascal Mueller and Bo Zhao

Four more keynotes were given on the second day. The first featured **Pascal Mueller** of Esri's development center in Zürich, who talked about Esri's efforts in the digital twin space. Four scales guide Esri's strategies: global, with the Earth regarded as a solid figure; regional, with emphasis on environmental issues; city, with detailed visualization of its 3D structures; and building, with links to the BIM standards. At each of these scales the approach recognizes the vital role played by humans. However, Mueller particularly noted the role of humans at the city scale where the emphasis is on approaching the city as it would be perceived by humans. This emphasis aligns with Esri's sensitivity to what it calls the "tech-first trap" where all emphasis is on technology. Pascal showed a video featuring the Google Geospatial API, which he anticipates will greatly enhance the power of urban-scale twins.

There are currently four major research areas in Pascal's team: generative AI, which is artificial intelligence geared to the creation of content; Simulation as a Service (SaaS); participatory planning; and immersive virtual and augmented reality. He sees the urban planning application, and associated scale, as currently driving Esri's investment in digital twins. Visualization has been advancing rapidly, as a means of creating photo-realistic images of cities, and as a set of tools for integrating schematics of features such as underground infrastructure into those images. He showed several very compelling examples, including the digital twin that the City of Gothenburg is developing in Sweden. A video was used to demonstrate the potential of virtual reality in planning, and the possibility of supporting the sharing of comments among stakeholders connected via the Cloud.

Because of the relatively late hour in Europe, questions were invited from the participants immediately after the presentation rather than waiting for the second keynote of the pair. The question of equity of access was raised: how might these digital twin technologies be made accessible beyond the digital divide, and to comparatively underresourced communities. In response, Pascal noted that Open Street Map is an enormously valuable and yet free resource, and that the heaviest computing load in the technology he had demonstrated was the conversion of scan data to the building mesh, a task that is often undertaken by contractors who enjoy massive economies of scale.

Following this short period of discussion, **Bo Zhao** began his keynote presentation. He began by reflecting on the transition between a Newtonian perspective that sees the world as a large machine governed by immutable laws, and the view advanced by Whitehead and others of the world as an organism in which humans play a central role. From this latter perspective he saw four major types of GIS. First, embodiment GIS is exemplified by wearable devices, where the GIS becomes an extension of the human. Second, hermeneutic GIS is the traditional form in which the geographic world is captured in a digital system and provides services that are interpreted by human users. Third, an autonomous GIS might include the GIS elements of autonomous vehicles and other field robots. And fourth, a background GIS is part of a place, not necessarily apparent to its users, and exemplified by the services of a smart city.

Of particular relevance to hermeneutic GIS is the growing possibility of fakes. Tools such as generative AI are being used to produce fake maps and fake images, though the history of imaginary geographies dates back centuries. Fakes can present significant threats to national security, and can be part of a growing industry of misinformation and disinformation. We are increasingly exposed to images of sunsets, landscapes, or advertising materials that have been doctored to enhance colors and other features. Some progress has been made in detecting fakes through analysis of colors and through violations of known geographic principles. But these concerns extend well beyond images to the spoofing of locations or email addresses by aggressive telemarketers.

Bo ended his presentation with some take-home messages. A humanistic approach is critical in the interactions between people, digital twins, and places, a point that echoes the earlier presentation by Pascal; a more-than-human approach can be helpful in considering the ethical and political implications of GIS and digital twins, and needs to be explored in depth. In the discussion, Bo recommended that in light of all this we should first and foremost be humble; Shaowen agreed but argued that in addition we should be bold.

Keynote Pairing IV – Clancy Wilmott and Trisalyn Nelson

The final pair of keynotes later on Tuesday morning was led off by **Clancy Wilmott** of UC Berkeley. She began with some fundamental questions about the scope and context of digital twins. The concept of a digital twin first came to the fore in aeronautical engineering and other forms of manufacturing, so she wondered what such systems might have in common with drilling platforms, conservation areas, or cities. How should we select the features of a system to represent and incorporate when building a digital twin, and what does the term “twin” imply in terms of equivalence? She took us on a virtual trip to the town of Bathurst in New South Wales to provide a concrete example, and asked us to imagine building a digital twin of this medium-sized town. As with all such towns it has a diverse human population and a diversity of programs and features, some of which are distinctly local. Some features, such as electric vehicle charging stations, may be increasingly relevant in the future, but other features such as playground equipment are of immediate importance to families with young children. The town includes highly educated people as well as discarded sharps. How will an effort to build a digital twin be organized and governed, and to what extent will its contents be driven by what is already mapped? She gave examples of local conditions that will pose problems for any solution that is imported from elsewhere, such as the difficulty of programming autonomous vehicles to detect and avoid collisions with kangaroos, and the problems caused by Australia’s high rate of continental drift, which requires that any high-definition mapping be regularly updated.

The final keynote was given by **Trisalyn Nelson** of UC Santa Barbara. Digital twins are built from knowledge about processes, which we are sometimes able to deduce from patterns in the classic cross-sectional approach of spatial analysis. But the principle of equifinality teaches that one pattern can be the result of numerous processes. Her efforts in Alberta over several years as part of a team studying Grizzly bears led to something close to a digital twin, and also to the training of highly qualified personnel. She argued that building a digital twin can be likened to an endless journey during which more and more information and new scientific knowledge is acquired and accumulated, but the end of a perfect representation is always unattainable. Many data issues arise during the journey, including issues of uncertainty, provenance, and ethical access, along with the what-if? scenarios that have long been part of geodesign.

Trisalyn asked whether a digital twin can be generic, or must always be tied to a specific use case. Use cases will determine the sets of data and process representations that are needed, but they will also determine whether an imperfect digital twin is fit for purpose.

In the discussion which followed, Clancy raised an argument that is often given in relation to autonomous vehicles: humans are flawed but machines can be entirely predictable. Yet a predictable machine would likely not anticipate the possibility of a black swan, or the impact of a tsunami from Japan on the Bay Area. Shaowen argued that workforce development will be a major issue in the context of digital twins, and required some fundamental thinking about what knowledge would need to be acquired by the next generation. Clancy wondered why cities are so often shown in video games as ruined: was this perhaps because ruined cities contain fewer people?

Lightning talks

In addition to the eight keynotes summarized above, the workshop also contained a series of lightning talks. These five-minute talks allowed many of the participants in the workshop to express their ideas about digital twins. **Sean Ahearn** of Hunter College gave the first talk and opened with a comment from a seat neighbor on the flight out from New York: were digital twins a little like teenage sex, with many people talking about it but few knowing what it is and even fewer doing it? He described efforts to build a digital twin of the city of New York and the various stages along the way since 2001. Many aspects of underground infrastructure are included, and Sean discussed some of the sources of uncertainty and efforts at quality assurance. **Guofeng Cao** of the University of Colorado, Boulder, focused on the modeling of uncertainty in digital twins. While progress has been made in recent decades in the measurement, modeling, and simulation of uncertainty in geospatial data sets, the need in digital twins to integrate multiple sources of data leads to a much more challenging problem. Moreover, there has been little attention to uncertainty in GeoAI. He showed some recent work that cascades uncertainties from multiple layers of data. He also discussed downscaling, which will likely be required when the spatial resolution of the layers in a digital twin is much finer than that of some of the original data.

Somayeh Dodge of UC Santa Barbara focused her talk on a digital twin for human–wildlife interaction. The pause in human activity caused by the recent pandemic, sometimes termed the anthropause, had an observable effect on wildlife behavior. She talked about Movebank, a worldwide repository of animal movement data, and compared it to data on the tracking of human movement. Bringing these two data sources together might lead to some interesting inferences, but would also raise issues with respect to spatial and temporal resolution. **Song Gao** of the University of Wisconsin Madison followed with a presentation about preserving privacy in mobility research. Location privacy can be protected by aggregating or geomasking locations, but Song talked about new research within GeoAI that attempts to protect privacy by using federated techniques that avoiding any sharing of individual locations.

Jeff Hamerlinck of the University of Wyoming talked about the emphasis on urban areas in digital twin discussions, and asked how some of that discussion might be adapted to rural areas. The urban/rural distinction is often oversimplified, but the most important issue might be what could be termed the "smart divide," that is, the distinction between areas that are dominated by highly educated populations and the technology industry on the one hand, and traditional economies on the other. He talked about advances in digital technology for agriculture, and about the concept of the smart village.

Junghwan Kim of Virginia Tech talked about the StreetView images captured by Google. Although these images of large cities have been the subject of much research, there has been much less on small and medium-sized cities. Analysis of 45 such cities shows great variation in spatial and temporal coverage, suggesting that efforts to build digital twins for small and medium-size cities may be frustrated. He suggested that this gap could be closed through the use of less expensive technology by cities and metropolitan planning organizations, but individual privacy may become an issue. **Mark Reynolds** of The Nature Conservancy closed the first batch of lightning talks by describing the efforts of TNC to adopt digital-twin methods in its conservation efforts at the Dangermond Preserve, providing an extended introduction to the information and concepts discussed by Kelly Easterday in her keynote.

Xiaojiang Li of Temple University asked how digital twins might help solve the challenges of urban heat islands. The models being used to simulate heat islands rely on enormous volumes of data at fine spatial resolutions, making it impossible to use such models in real-time management. Thus, his talk focused on the use of GPUs (graphic processing units) to speed up the simulation process. Speedup of factors of 400,000 can be readily achieved. He demonstrated Web-based simulations for several US cities using his HeatExpo site. **Ruopu Li** of Southern Illinois University at Carbondale built on Jeff Hamerlinck's earlier talk about working beyond the "smart divide." His work is grounded in the theoretical foundation of social-technical systems, and focuses on the comparison of two cities in southern Illinois: Cairo and Carbondale. He defines a social-technical digital twin as "a digital replica that simulates the process of communities engaging with infrastructures/technologies for specific purposes."

David Page of Oak Ridge National Laboratory argued that traditional geospatial workflows are aimed at a product and thus tend to be stove-piped, whereas the purpose of a digital twin may be much more generic. He used the digital surface model being built at ORNL as an example, arguing that the real value of elevation data obtained from LiDAR lies in the point cloud, and the ability to rapidly produce a range of products from it provides downstream encapsulations of that value. Thus, a digital twin can be seen as a form of archive or an infrastructure for data and model management, supporting a range of projects. **Shashank Karki**, a masters candidate from Virginia Tech, returned the conversation to Pascal Mueller's discussion of the building scale. Shashank discussed building digital twins of the interior of buildings using tools such as LiDAR and SLAM (simultaneous location and mapping) that are effective ways to create interior models. These models are however static; this project is aimed at including people and their movements to incorporate additional dynamism. Initially movements were obtained from accelerometers and cameras, but these data are being combined with the results of interviews and real-time LiDAR. It is hoped that the kinds of information being provided by this project will be used to improve the layout of buildings and the ability of building managers to respond to emergencies.

In the final lightning talk **Damian Spangrud** of Esri provided an industry perspective on digital twins. He began by suggesting that digital twins were a logical extension of the role that has always been played by maps, as attempts to represent important elements of the geographic world and support decision-making. But digital twins take advantage of major developments in technology, including remote sensing, LiDAR, and sensor networks, and they build on knowledge of relationships and behaviors. He suggested that in future digital twins would push further in data integration, would integrate smoothly with the concept of the metaverse, would benefit further from new real-time feeds of data, and would allow the user to move smoothly between indoors and outdoors.

A digital twin of the ASU map collection

The final activity of the first day of the workshop, prior to a reception and dinner, was led by **Matt Toro**, the head of the ASU library's map collection, and included an online demonstration and then a tour and discussion. Matt and his team have constructed a digital twin of the space occupied by the map collection in the Hayden Library. It includes a three-dimensional representation of the space, including its furniture, and allows the user to search for specific maps in the collection and see their locations in the map drawers. The scale is centimeters, which is sufficient to provide compelling visualizations. Digital scans of many of the maps have been integrated with their catalog records, and in future it will be possible to superimpose the maps on visualizations of the surface of the Earth.

Synthesis

By the middle of the second day many ideas had been presented, and many suggestions had been made about research needs and the future directions of digital twins. This section attempts to synthesize, by addressing several issues that emerged repeatedly during the workshop.

Definitions

Although authors have found many ways of expressing the meaning of the term digital twin, there is in fact little evidence of disagreement. As noted earlier, a digital twin is "A virtual representation of the real world, including physical objects, processes, relationships, and behaviors" (Esri). Some definitions include the purpose of the digital twin, such as the evaluation of what-if scenarios, and some go so far as to claim that the representation is perfect, and that all results obtained from the digital twin will exactly mirror results that might be obtained from the reality represented by the digital twin—in effect, that the digital twin passes the Turing test.

Metaphors

As is so often the case, when definitions are unclear we turn to metaphors. A number of metaphors for the building and use of digital twins surfaced during the discussions, and might provide simple ways of characterizing their use and importance. Trisalyn suggested that we should think of building a digital twin as a journey, with benefits that occur somewhat continuously. Digital twins were also likened to a moonshot, and to President Kennedy's 1961 commitment to landing humans on the moon by the end of the decade. Mike Goodchild related how in 1998 he had suggested during a Congressional Briefing that the construction of a Digital Earth, as defined earlier that year by Al Gore, might be a suitable moonshot with a target date of 2005, and arguably Google's release of Google Earth in 2005 was a fair approximation to what Gore had in mind. Several of the digital twins discussed during the workshop seemed to justify an umbrella metaphor, as an infrastructure for aggregating and preserving the data and models used in a collection of research projects. Such an infrastructure would ensure that future projects could easily leverage the results of prior projects, and the umbrella metaphor might suggest ways of fleshing out the contents and stimulating new research ideas.

Objectives

Many distinct types of objectives were suggested and discussed during the workshop. The work of Pascal's group in Zürich is strongly motivated by urban planning and design, and the ability to visualize and model the results of decisions about architecture, transportation, and service infrastructure. In this context digital twins might be seen as reflective of earlier discussions about agent-based models, as ways of integrating knowledge of pattern and process into a form that is readily usable in a decision-making context. Visualizations and dashboards provide easy ways of seeing the implications of adjustments to complex systems, and ways of presenting planning decisions as stories.

It is harder, however, to see how digital twins might themselves result in new knowledge. It is often argued that the value of agent-based models to science lie in the possibility of emergent properties, that is, properties that would become apparent to the researcher only after the agent-based model or digital twin is built. Suppose for a moment that some new and unexpected phenomenon or behavior does in fact emerge from a digital twin—a behavior or phenomenon that has never been observed in reality. In this situation the researcher must somehow determine whether the new phenomenon is real, and not an artifact of the imperfect digital twin and the uncertainty that is certainly present in its representations of the real world and its processes. If policies are based on the new phenomenon, they will necessarily involve risk. In short, it is hard to see digital twins as producers of new scientific knowledge.

That does not necessarily mean, however, that digital twins are of no interest to science. The emergence of new phenomena in a digital twin may stimulate research to improve the data and models underlying the digital twin, or to fill the missing pieces in a digital-twin umbrella. Thus, a digital twin can be an invaluable component of a scientific infrastructure, but as a consumer and motivator rather than as a producer. That position is consistent with the ways in which digital twins have been discussed during the workshop; no participant argued or provided evidence that digital twins can advance scientific knowledge in and of themselves, or pointed to actual discoveries made by digital twins.

Reliability and validity

As a representation, a digital twin must necessarily be incomplete, since it is impossible to create a perfect digital representation of any part of reality, and impossible therefore to create a perfect digital twin—"the map is not the territory" (Korszybski, 1933). It follows that the representations in the digital twin are subject to uncertainty, and that these uncertainties propagate through to its predictions and visualizations. In this respect digital twins are no different from any GIS, or for that matter from views of the world as obtained through the human eye. The problem has been dealt with most often by arguing fitness for purpose: that it is the use case that must determine the levels of uncertainty that can be accepted. For example, the spatial resolution of the data in a digital twin must be sufficiently fine to exclude any possibility that the desired predictions will be significantly impacted by real events, features, or relationships that are too small to have been captured. In short, there is an essential and unavoidable relationship between a digital twin and its associated use cases.

It is important that appropriate terms be used in discussions of the uncertainties associated with digital twins. Accuracy is rarely appropriate, since it implies the existence of a true value that has been distorted by the process of measurement. Precision is also inappropriate if used in the sense of the degree of detail used to report a measurement, since it is in this sense not a property of the measurement but only of its reporting. Validity and validation are problematic because they imply a binary answer—yes the results are valid or no they are not—thus hiding the fact that no results from a digital twin can ever be perfect.

Much is now known about uncertainty in geospatial data as a result of more than three decades of research. However, the use of digital twins and advances in geospatial data acquisition raise new issues. It became clear from several presentations that major research questions need to be addressed in digital-twin research:

- How can the uncertainties in the new data sources (e.g., social media sources, LiDAR acqusition of 3D city structure) being exploited in digital twins be measured, represented, modeled, and propagated through digital twin simulations?
- What research is needed into the ways in which independent uncertainties from multiple data sources combine in the simulations of digital twins? Are there limits to the degree to which multiple data sources can be fused?
- Are there ways of introducing uncertainty into the photo-realistic visualizations of digital twins that can be used to understand and assess the usefulness and usability of the models?
- How can users be alerted to the importance of uncertainty and the need to propagate it into digital twin predictions?
- What are the consequences of ignoring uncertainty in working with digital twins? How does this uncertainty impact real-world decisions that might be made based on a digital twin?

Technical issues of digital twin design

As digital twins proliferate, increasing attention is being devoted to their interactions. The buildingSMART International organization has published a paper that addresses the concept of a digital twin ecosystem as a collaboration between interacting digital twins, and while the DITTO Project is actively advancing the notion of a digital twin for the oceans, it will clearly need to address relationships between the oceans and the terrestrial third of the globe, if only in coastal zones. Digital twins may coexist in the same geography, perhaps at different scales, raising more questions about the technology of interactions.

At the workshop there was discussion of the problems that occur when predictions are available from multiple digital twins. Parallels were drawn with modeling global climate change, since there are now tens of models available, each producing predictions and often without explicit attention to uncertainty. It has become common to refer to these multiple predictions as ensembles. However we do not know how the individual predictions sample the range of possible predictions, or how they relate to the reality that will eventually emerge. There is no justification, for example, for assuming that the mean of an ensemble is more valuable than any of the individual members, or that reality will stay within the range of the ensemble. Similar arguments can be made about the predictions of digital twins.

One of the more important technical issues concerns the nature of the digital twin's user interface, and the user experience. What design of user interface is best suited to digital twins that are intended for decision making, or for advancing science, or for predicting the results of what-if scenarios? What can be learned from the design of video games, from participatory and collaborative GIS, and from geodesign?

If a digital twin is a repository of data, models, and tools, as suggested by Kelly Easterday's use of the term, and a way of integrating the materials used by numerous research projects into a valuable collective resource, then a number of questions arise to which we do not necessarily have the answers. How does one structure such a repository, and what tools will be needed to search it? Should all data sets be reduced to a common geographic representation, or left in their original form and restructured on the fly as needed? How should uncertainty be captured, represented, and propagated? All of these issues are fundamental to GIScience.

Ethical and societal issues

A recent white paper from the GeoEthics Project, "Locational Information and the Public Interest," has drawn attention to the rapidly expanding set of ethical issues raised by the extensive use of geospatial technology. Many of these issues are directly relevant to digital twins, including the potential for surveillance and invasion of privacy and the potential for drawing inappropriate inferences from geospatial analytics. There is a need for reflection, not only on these issues and others in relation to digital twins, but also on the broader issues of digital twins for society. In this respect the research agenda should seek to advance the established literature on the societal impacts of GIS (e.g., Pickles, 1995). Following are some possible topics for research and further discussion:

- Can digital twins be repurposed from their intended applications, and might such repurposing raise ethical issues? How might the user interface of digital twins be designed to minimize such possibilities?
- When individuals are represented and modeled in a digital twin, how might their privacy be protected? How appropriate are the methods described in the workshop by Song Gao?
- How should the results, predictions, and visualizations of a digital twin be presented while placing appropriate emphasis on uncertainty?
- What rules and regulations currently govern the ownership of digital twins and their data and processes? Is there a need for additional regulation?
- What criteria should a project be required to satisfy to merit the designation as a digital twin? Is there evidence that the term is being misused, or that false claims are being made about specific instances?

- What levels of openness and transparency should be required of digital twins when used by a) public agencies, b) research projects, c) private consulting companies, or d) commercial software producers?
- In what ways might the visualization functions of digital twins be engineered to mislead?
- How should organizations developing digital twins incorporate potential positive and negative consequences of their efforts into decision-making about the design of those twins? Should certain data stream or functions be excluded, who should makes these decisions and how?

The very use of the term “twin” might be seen as raising a question of ethics, since it implies that results from the twin will be identical to those obtained by querying the real world—in other words, that the twin passes the Turing test. Yet developers and marketers of digital twins will be aware that uncertainties will always be present in the twin’s data and in its simulation models, and that processes such as resampling will have added to those uncertainties. Thus by using the term the developers and marketers may be perpetuating a potentially dangerous misunderstanding. Perhaps we should encourage the replacement of “twin” in environmental and social contexts with a less definitive term: “mirror” or “doppelganger” might be suggested. On the other hand even the biological use of the term does not imply exact equivalence.

Education and training

Considerable interest was expressed at the specialist meeting in the development of an agenda for education and training. At this time there is a general lack of publications, textbooks, and teaching materials on digital twins, and little information is available to help an instructor to introduce digital-twin content into courses. This is similar to the situation facing GIS instruction in the 1980s, and it may be helpful to review the efforts made at that time in response. Following are some potential topics for follow-on activities:

- What principles underlie digital-twin technology that are likely to survive the next decade and might form the basis of an upper-division course?
- How should future users and developers of digital twins be trained, on what existing software?
- What example applications of digital twins might form a suitable gallery for a course? Have some digital-twin projects failed, and might these failures be included in a course?
- Should digital twins be taught as one or more separate courses, or should they be micro-inserted into existing courses?
- What understanding of the principles of spatial uncertainty, geovisualization, and ethics should be a) expected of enrollees in digital-twin courses, or b) taught as part of those courses?
- Must training in digital-twins span disciplines? If so, what disciplines should participate and how should that participation be balanced?

Concluding comments

The workshop was designed, and keynote speakers and participants were selected, in order to open the growing interest in digital twins to the open discussion, critical reflection, and agenda-setting that are typical of academia. Digital twins are clearly an important topic for agencies responsible for environmental management, for software houses supporting urban planning, and for companies providing software to manufacturers. Yet they raise some profound issues, including the role of uncertainty in fields such as social and environmental science where uncertainty is inescapable. Thus it is important, if not essential, for the academic community to do what it does best: to reflect on progress to date and on likely future directions.

This role is what has guided the design and execution of this workshop, and the contributions of the keynote speakers and participants. The organizers hope that the online recordings and slides and this report will help to advance the technology and practice of digital twins, and encourage educators to introduce the principles of digital twins into their classrooms and online courses.

References

Guo, H., M.F. Goodchild, and A. Annoni, editors, 2020. *Manual of Digital Earth*. Singapore: Springer Open.

Korzybski, A., 1933. *Science and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics*. Lakeville CT: International Non-Aristotelian Publishing Co.

Pickles, J., editor, 1995. *Ground Truth: The Social Implications of Geographic Information Systems*. New York: Guilford.